CLAIMS:

1. A computer-implemented method for hashing a body of text, the method comprising:

obtaining a body of text;

deriving a hash value representative of content of the body of text, perceptually distinct bodies of text having hash values that are substantially independent of each other.

2. A method as recited in claim 1, wherein perceptually distinct bodies of text have hash values that are independent of each other.

3. A method as recited in claim 1 further comprising comparing hash values of two bodies of text to determine if such values match.

4. A method as recited in claim 1 further comprising comparing hash values of two bodies of text to determine if such values substantially match.

5. A method as recited in claim 4 further comprising indicating whether such values substantially match.

6. A computer comprising one or more computer-readable media having computer-executable instructions that, when executed by the computer, perform the method as recited in claim 1.

7. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method as recited in claim 1.

8. A method for facilitating recognition of content of a body of text, the method comprising:

filtering the content a body of text to remove elements of the content;

determining a recognition representation of the content of such body based upon the filtered subtext.

9. A method as recited in claim 8, wherein the filtering is text-sifting.

10. A method as recited in claim 8, wherein the determining comprises calculating the recognition representation as a hash value that identifies the content in the body.

11. A method as recited in claim 8, wherein the determining comprises calculating the recognition representation as a hash value that is proximally similar to other bodies of text having similar semantic content.

12. A method as recited in claim 8, wherein the filtering comprises removing superfluous elements from the content of the body.

13. A computer comprising one or more computer-readable media having computer-executable instructions that, when executed by the computer, perform the method as recited in claim 8.

14. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method as recited in claim 8.

15. A computer-implemented method for hashing a body of text, the method comprising:

obtaining a body of text;

deriving a hash value representative of the body of text, perceptually similar bodies of text having proximally similar hash values.

16. A method as recited in claim 15 further comprising comparing hash value of a body of text to determine if such value is proximally near hash values of a group of bodies of text having proximally clustered hash values.

17. A method as recited in claim 16 further comprising grouping the body of text with the group of bodies of text if the hash value of such body is proximally near the values of the group.

18. A computer comprising one or more computer-readable media having computer-executable instructions that, when executed by the computer, perform the method as recited in claim 16.

19. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method as recited in claim 16.

20. A method for facilitating recognition of content of a body of text, the method comprising:

obtaining a body of text;

determining a self-synchronized recognition representation of the content of such body.

21. A method as recited in claim 20, wherein the self-synchronized recognition representation is derived from a subset of the content of the body of text.

22. A method as recited in claim 20, wherein the self-synchronized recognition representation is derived from a subset of the content of the body of text, the subset excludes superfluous elements of the content of the body of text.

23. A method as recited in claim 20, wherein the self-synchronized recognition representation is derived from a subset of the content of the body of text.

24. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method as recited in claim 20.

25. A computer comprising one or more computer-readable media having computer-executable instructions that, when executed by the computer, perform the method as recited in claim 20.

26. A method for facilitating recognition of content of a body of text, the method comprising:

filtering the content of a body of text to select a subset of content of such body;

determining a recognition representation of the content of such body based upon the selected subtext.

27. A method as recited in claim 26, wherein the filtering is text-sifting.

28. A method as recited in claim 26 further comprising storing the recognition representation in a database, the recognition representation being associated with the body of text from which it was determined.

29. A method as recited in claim 26, wherein the determining comprises calculating the recognition representation as a hash value that identifies the content in the body.

30. A method as recited in claim 26, wherein the determining comprises calculating the recognition representation as a hash value that is proximally similar to other bodies of text having similar semantic content.

31. A method as recited in claim 26, wherein the filtering comprises removing elements from the content of the body.

32. A method as recited in claim 26, wherein the filtering comprises removing superfluous elements from the content of the body.

33. A method as recited in claim 26, wherein the filtering comprises removing elements from the content of the body, wherein at least some of the elements removed are associated with a format of the content of the body.

34. A method as recited in claim 31, wherein the removing comprises:

converting white space in the body of text into single spaces;

purging all content of the body of text that is not letters or spaces;

converting all content of the body of text into one form of capitalization.

35. A method as recited in claim 31, wherein the removing comprises:

referencing a list of common words;

purging all words from the body of text that are on the list of common words.

36. A method as recited in claim 26, wherein the filtering comprises cryptographically extracting the subset of text of such body.

37. A method as recited in claim 26, wherein the subset has a fixed size that is independent of size of the subset's body of text.

38. A method as recited in claim 26, wherein the subset has a variable size that is dependent upon size of the subset's body of text.

39. A method as recited in claim 26, wherein the filtering comprises:

removing superfluous elements from the content of the body to produce filtered text;

cryptographically extracting the subset of text of such body from the filtered text.

40. A method as recited in claim 26 further comprising comparing recognition representations of text of at least two bodies of text.

41. A method as recited in claim 40 further comprising indicating a match if recognition representations of text of at least two bodies of text substantially match.

42. A method as recited in claim 26 further comprising:

comparing recognition representation of text of a body of text with recognition representations of text of a group of bodies;

grouping the body with the group if all compared recognition representations are proximally similar.

43. A computer comprising one or more computer-readable media having computer-executable instructions that, when executed by the computer, perform the method as recited in claim 26.

44. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method as recited in claim 26.

45. A method for facilitating detection of textual similarity, the method comprising:

comparing recognition representations of text of at least two bodies of text, wherein such recognition representations are computed by:

text sifting text of the bodies of text to select a subset of text for each body;

determining such recognition representation of the text for each body based upon the selected subtext of each body;

indicating a match if recognition representations of the text of at least two of the bodies substantially match.

46. A method as recited in claim 45, wherein the determining comprises calculating the recognition representation as a hash value that identifies the content of the body.

47. A method as recited in claim 45, wherein the text sifting comprises cryptographically extracting the subset of text of such body.

48. A method as recited in claim 45, wherein the text sifting comprises:

removing superfluous elements from the text of a body to produce filtered text;

cryptographically extracting the subset of text of such body from the filtered text.

49. A computer comprising one or more computer-readable media having computer-executable instructions that, when executed by the computer, perform the method as recited in claim 45.

50. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method as recited in claim 45.

51. A method of manipulating content of a source body of text, the method comprising:

obtaining a source body of text;

generating content of a target body of text by deriving the content of the target body from the source body;

wherein the content of the target body has a self-synchronized recognition representation that does not substantially match a self-synchronized recognition representation of the content of the source body.

52. A method as recited in claim 51, wherein the content of the target body has a self-synchronized recognition representation that does not match a self-synchronized recognition representation of the content of the source body.

53. A method as recited in claim 51, wherein the self-synchronized recognition representations are determined by producing a hash value of a subset of the content of a body, wherein the subset excludes superfluous elements.

54. A text recognition system, comprising:

text retriever for obtaining body of text;

text sifter for selecting a subset of text of such body;

recognition representation determiner for determining a recognition representation of the text of such body based upon the selected subtext.

55. A system as recited in claim 54 further comprising a database for storing the recognition representation in association with the body of text from which it was determined.

56. A system as recited in claim 54, wherein the determiner comprises a calculator to calculate the recognition representation as a hash value that identifies the content of the body.

57. A system as recited in claim 54, wherein the determiner comprises a calculator to calculate the recognition representation as a hash value that is proximally similar to other bodies of text having similar semantic content.

58. A system as recited in claim 54, wherein the text sifter comprises a extractor for cryptographically extracting the subset of text of such body.

59. A system as recited in claim 54 further comprising a comparator for comparing recognition representations of text of at least two bodies of text.

60. A system as recited in claim 54 further comprising:

a comparator for comparing recognition representations of text of at least two bodies of text;

an indicator for indicating a match if recognition representations of text of at least two bodies of text substantially match.

61. A system as recited in claim 54 further comprising:

a comparator for comparing recognition representation of text of a body of text with recognition representations of text of a group of bodies;

a categorizer for grouping the body with the group if all compared recognition representations are proximally similar.

62. A computer-readable medium having stored thereon a data structure, comprising an library containing bodies of text where at least one body is associated with a recognition representation determined by the system as recited in claim 54.

63. A computer-readable medium having stored thereon a data structure, comprising:

a first data field containing a body of text;

a second data field derived from the first field by text sifting the text of such body to select a subset of text of such body and determining a recognition representation of the text of such body based upon the selected subtext;

a third data field functioning to delimit the end of the data structure.

64. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method comprising:

obtaining a body of text;

deriving a hash value representative of content of the body of text, perceptually distinct bodies of text having hash values that are substantially independent of each other.

65. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method comprising:

obtaining a body of text;

deriving a hash value representative of the body of text, perceptually similar bodies of text having proximally similar hash values.

66. A computer-readable medium having computer-executable instructions that, when executed by a computer, performs the method comprising:

obtaining a body of text;

text sifting the text of such body to select a subset of text of such body;

determining a recognition representation of the text of such body based upon the selected subtext.